TECHNIQUES FOR MOLECULAR ANALYSIS

Tackling the plant proteome: practical approaches, hurdles and experimental tools

Jocelyn K.C. Rose[1,*], Sajid Bashir[2], James J. Giovannoni[3], Molly M. Jahn[4] and Ramu Subramanian Saravanan[1]

[1]*Department of Plant Biology, Cornell University, Ithaca, NY 147853, USA,*

[2]*Department of Chemistry, Texas A & M University, Kingsville, TX 78363, USA,*

[3]*USDA-ARS Plant, Soil and Nutrition Laboratory, Ithaca, NY 147853, USA, and*

[4]*Department of Plant Breeding, Cornell University, Ithaca, NY 147853, USA*

Summary

The study of complex biological questions through comparative proteomics is becoming increasingly attractive to plant biologists as the rapidly expanding plant genomic and expressed sequence tag databases provide improved opportunities for protein identification. This review focuses on practical issues associated with comparative proteomic analysis, including the challenges of effective protein extraction and separation from plant tissues, the pros and cons of two-dimensional gel-based analysis and the problems of identifying proteins from species that are not recognized models for functional genomic studies. Specific points are illustrated using data from an ongoing study of the tomato and pepper fruit proteomes.

Keywords: protein, proteomics, protein extraction, 2-D gel electrophoresis, liquid chromatography, mass spectrometry, peptide mass fingerprinting.

Introduction

The last couple of years have witnessed an explosion of activity in the application of genome-scale gene expression profiling tools to explore biological systems. While not hypothesis driven, the ability to qualitatively and quantitatively monitor mRNA and protein populations raises the tantalizing prospect of deciphering the functional and regulatory networks that represent the bridge between genotype and phenotype. This in turn has catalyzed the 'systems biology' paradigm (Ge *et al.*, 2003; Patterson and Aebersold, 2003 and a special issue of *Plant Physiol.* **132**, 2003). In practical terms, this can take the form of global transcript profiling and a prodigious amount of information is being generated by cDNA and oligonucleotide-based microarray studies in a wide variety of species. This technology has matured to the point where it has become readily available to the community of plant biologists and microarray analyses are included in, or form the basis of, a rapidly growing number of papers in a broadening spectrum of plant biology journals. The transcriptomic perspective is particularly attractive as RNA is chemically homogeneous and is relatively easy to extract, manipulate *in vitro*, amplify and sequence. This allows high-throughput, parallel, quantitative analysis of many thousands of distinct and well-defined gene products, representing a comprehensive coverage of the transcriptome (see Alba *et al.*, 2004).

An alternative strategy is to target the next molecular link in the information chain from gene to phenotype and evaluate the proteome; examining the protein status of a cell type, tissue, organ or whole organism. Proteomics is an increasingly ambiguous term that is now being applied to almost any aspect of protein expression, structure or function. Indeed, numerous recent reviews have highlighted the remarkable developments in diverse areas of protein science that fall within the proteomics arena, including protein structure, function and protein–protein, or protein–ligand interactions (e.g. Cho *et al.*, 2004; Jung and Lee, 2004; Mann and Jensen, 2003; Patterson and Aebersold, 2003; Schmid, 2002; Yanagida, 2002; Yarmush and Jayaraman,

2002; Zhu *et al.*, 2003). However, as defined here it relates to the systematic study of the protein complement of the genome and the focus of this review will rest on experimental approaches that are used for comprehensive protein expression analysis; analogous to transcript profiling.

Why proteomics?

In contrast to the relative ease of mRNA extraction and experimental manipulation that are required with a transcriptomic analysis, proteins present numerous challenges: they are physicochemically highly heterogeneous and structurally complex which dramatically complicates their extraction, solubilization, handling, separation and identification and no technology currently exists that is equivalent to PCR to amplify low abundance proteins. As will be described later, some of these difficulties are particularly severe when working with plant tissues. Perhaps the first question that should be asked then is whether an assessment of global gene expression should be attempted at the protein level if other potentially simpler approaches are available? Indeed, the enormous power and cost-benefits of DNA chip-based studies ensure that they are likely to provide the predominant impetus for genome-scale studies for the foreseeable future. In many cases then microarray analysis may represent a more appropriate strategy, particularly as a first step. However, reliance on this technique as the sole tool for profiling gene expression has a number of limitations. Perhaps the most important of these is that mRNA levels appear not to be a consistent indication of cognate protein abundance and several studies have revealed poor correlations between changes in the abundance of specific mRNAs and their corresponding proteins (Anderson and Anderson, 1998; Anderson and Seilhamer, 1997; Gygi *et al.*, 1999a; Ideker *et al.*, 2001). Such studies, when considered alongside the mathematical demonstration that an understanding of even relatively simple gene networks requires knowledge of both the underlying protein and mRNA expression levels (Hatzimanikatis and Lee, 1999a), clearly suggest that the full value of transcriptomics will only be realized when coupled with proteomic studies.

A related consideration is that substantial regulation of cellular events can occur at the protein level with no apparent changes in mRNA abundance. Post-translational modification of proteins can result in a dramatic increase in protein complexity without a concomitant increase in gene expression. Indeed, while one yeast gene on average encodes between one and three distinct modified proteins, in humans this number is typically between three and six and in some cases substantially higher (Wilkins *et al.*, 1996). Consequently, the human genome, which contains approximately 30 000 genes (Lander *et al.*, 2001; Venter *et al.*, 2001), has been estimated to encode anywhere from 200 000 to 2 million proteins (Service, 2001). It is

reasonable to assume that plant proteomes will be of a similar order of magnitude.

Additional arguments that are typically forwarded in favor of using proteomics as an experimental platform are based on the notion that as the proteome reflects the expression of the molecules that more directly influence cellular biochemistry, this provides a more accurate representation of cellular state than profiling the expression of mRNAs, which represent information intermediates. Similarly, considerable value can be placed on the ability to isolate subcellular protein fractions and thus gain insight into subcellular localization and even function, or to isolate multi-subunit protein complexes whose constituent polypeptides cannot be predicted from DNA sequences or mRNA abundance.

There are therefore many compelling reasons to consider undertaking comprehensive analyses of global protein expression and, while lagging behind pioneering studies in model prokaryotic and eukaryotic species (Ghaemmaghami *et al.*, 2003; Washburn and Yates, 2000), the term proteomics is now well established in the lexicon of plant biologists. The majority of plant proteomic studies to date, as defined earlier, can be divided into two basic categories. The first involves protein profiling of biological material with the aim of separating, sequencing and cataloging as many proteins as possible. Here, the objective is to establish the protein framework of a biological system, much as expressed sequence tag (EST) provides a snapshot of the transcript complement. However, as will be described later, entire proteomes of single cell types cannot yet be fully mapped, let alone those of complex tissues. A related and potentially complementary strategy is to target subcellular proteomes, thereby dramatically reducing the protein complexity of a particular extract and revealing important information regarding subcellular localization (Dreger, 2003). A number of studies are underway to survey the proteomes of a spectrum of plant subcellular compartments (Canovas *et al.*, 2004) including the chloroplast and constituent membranes (Kleffmann *et al.*, 2004; Lonosky *et al.*, 2004; Peltier *et al.*, 2000, 2002; Vener *et al.*, 2001), mitochondria (Bardel *et al.*, 2002; Kruft *et al.*, 2001; Millar *et al.*, 2001; Werhahn and Braun, 2002), endoplasmic reticulum (Maltman *et al.*, 2002), peroxisome (Fukao *et al.*, 2002) and amyloplast (Andon *et al.*, 2002). Several plant subproteome databases are now accessible online (see Tables 1 and 2). The second basic category of proteomic analysis can be termed comparative proteomics, where the objective is not to identify the entire suite of proteins in a particular sample, but rather to characterize differences between different protein populations. This approach is thus somewhat analogous to comparative DNA microarray profiling. Examples might include proteins from wild type versus mutant plants, or tissues at different developmental stages or following responses to external stimuli.

Table 1 Examples of online proteomics and mass spectrometry-related resources

Resource	URL
Protein analysis: techniques, tools, and problem-solving guidelines	http://www.techsupport.amershambiosciences.com/
	http://www.bio-rad.com
	http://www.invitrogen.com
	http://www.protocol-online.org
	http://www.weihenstephan.de/blm/deg/
	http://www.expasy.ch/ch2d/protocols
	http://www.genomicsolutions.com
	http://www.mgif.rserv.uga.edu/proteomics/sample_submission.html
	http://www.plantpath.unl.edu/llane/text/geltips.html
	http://www.aesociety.org/
	http://www.kendricklabs.com/
	http://www.lecb.ncifcrf.gov/flicker/
	http://www.bioanalyticaltech.com
Protein digestion and mass spectrometry sample preparation, analysis, and sequencing	http://www.spectroscopynow.com
	http://www.proteome.org.au/index.asp
	http://www.public.iastate.edu/~kamel/mstutorial.html
	http://www.matrixscience.com/
	http://www.ionsource.com
	http://www.functionalgenomicscenter.com/proteomics/protocols.cfm
	http://www.chem.cmu.edu/cma/links.html
	http://www.tucf.org/index.htm
	http://www.massspec.unm.edu/sample.htm#prepl
	http://www.albany.edu/genomics/proteomics-instrumentation.html
Proteomics and mass spectrometry courses, and training programs	http://www.genebio.com/products/products_proxemis.htm
	http://www.au.expasy.org/gpc/training/
	http://www.proteome.org.au/category.asp?category_id=41
Proteomic data mining tools: sequence prediction, and comparison	http://www.jura.ebi.ac.uk:8765/ext-
	http://www.genequiz//genequiz.html
	http://www.pedant.gsf.de/
	http://www.pat.sdsc.edu
	http://www.cbs.dtu.dk/services/SignalP/
	http://www.psort.nibb.ac.jp/
	http://www.cbs.dtu.dk/services/TMHMM/
MS Search engines	http://www.matrixscience.com/
	http://www.prowl.rockefeller.edu/
	http://www.prospector.ucsf.edu/
Proteomics journals and societies	http://www.wiley-vch.de/publish/en/journals/alphabeticIndex/2120/
	http://www.mcponline.org/
	http://www.pubs.acs.org/journals/jprobs/
	http://www.wiley-vch.de/publish/en/journals/alphabeticIndex/2027/
	http://www.bentham.org/cp/
	http://www.wiley.com/legacy/wileychi/genomics/cfg.html
	http://www.openmindjournals.com/genomics.html

A number of reviews specifically devoted to plant proteomics have been written over the last few years (e.g. Canovas *et al.*, 2004; Heazlewood and Millar, 2003; Kersten *et al.*, 2002; Roberts, 2002; Rossignol, 2001; van Wijk, 2001). These provide valuable and detailed insights into the breadth and scope of both protein profiling and comparative proteomic studies in plants. However, the original concept underlying the set of reviews in this journal issue was to provide a series of practical guides to researchers who are interested in using functional genomics tools to address a particular problem in plant biology, and yet who might not have the necessary background, budget or infrastructure to embark on a major large-scale genomics project. The potential usefulness of this type of review associated with proteomics is suggested by the anecdotal experiences of many plant proteomics laboratories who are frequently contacted with requests from colleagues across campus and beyond for assistance with protein-related projects. This can take many forms, from simple sequencing and identification of a single polypeptide or a protein complex, to more complex protein expression profiling a series of tissues. A growing number of plant researchers clearly wish to include the study of protein populations in their experimental toolbox, but without major investment in capital equipment. This review therefore does

not attempt to summarize the diversity of existing plant proteomics studies, nor to spotlight emerging technologies such as tandem affinity purification (TAP) tagging to isolate and identify protein complexes (Puig *et al.*, 2001; Rohila *et al.*, 2004). Rather, it is aimed at laboratory-based researchers who are interested in entering the field. An overview is provided of what might be expected when undertaking a plant proteomics study in terms of the conceptual and plant-specific technical challenges that are faced, and the types of data that are typically generated, particularly in contrast with other approaches to genomic-level gene expression profiling. Additional reviews that provide more detail of specific issues are cited where appropriate.

Primary considerations and experimental design

It would be impractical to attempt to summarize in this review the range of experimental techniques and types of data that can be generated through studies of complex protein populations. Moreover, some objectives such as the detailed cataloging of a particular subproteome, are currently specialist undertakings requiring substantial funding and in-house expertise. For example, in a *tour de force* of 'shotgun proteome analysis' Koller *et al.* (2002) described the identification and tissue-specific expression of more than 2500 unique proteins from several rice tissues. While this type of detailed surveying approach will certainly become more commonplace, current limitations in equipment availability, specialist knowledge, bioinformatics infrastructure and financial investment place such studies beyond the reach of many biologists. In addition, the data that are generated represent a static non-quantitative catalog of the protein architecture of a sample. In our experience, the most common type of request for assistance from plant biologists is with comparative proteomics projects, where the objective is to reveal changes in protein expression between samples, sometimes to complement parallel microarray analyses.

Although not the only options, some popular experimental approaches that are currently used to evaluate and contrast complex protein populations are summarized in Figure 1. The key elements include the extraction and detection of as many proteins as possible whilst minimizing post-extraction artifacts, accurate quantification and comparison of samples, protein identification and finally integration of protein expression information with other diverse data sets. Practical considerations and challenges associated with the various steps are discussed below; however, two overarching questions relate to the proportion of the entire proteome that can currently be analyzed by one or multiple techniques, and whether this represents a significant impediment to effective comparative proteomic analysis. Such questions are still difficult to answer, but in a recent review (Patterson, 2004) it was suggested that even

under optimal conditions, combining a range of separation and identification strategies, approximately 25% of the expected proteome might be observed at best. At the practical level even lower percentages may routinely be expected. Other related considerations are that certain subsets of proteins are extremely resistant to extraction, solubilization or subsequent separation, as described in the section below. Moreover, the large dynamic range of protein concentrations in eukaryotic cells, which is typically estimated as 10^5–10^6 (Patterson and Aebersold, 2003), means that low abundance proteins are typically not detected using the general approach that is shown in Figure 1, which would typically encompass a concentration range of 10^3 (Patterson, 2004). The first wave of identified proteins in most such proteome studies includes the major abundant 'housekeeping' proteins in fundamental metabolic pathways and not proteins that are expressed at low levels, such as most regulatory proteins, which are not generally detected in most proteomic studies. This problem is particularly marked in tissues with protein pools that are dominated with a few abundant polypeptide species, such as Rubisco (ribulose bisphosphate decarboxylase/oxygenase), which is the world's most abundant protein (Ellis, 1979) and can comprise more than half of total leaf protein in some species.

Current technologies therefore only provide a narrow window on the proteome, which substantially limits the ability to define and monitor the dynamic nature of protein networks. Some of the underlying technical limitations are outlined below, together with some possible strategies to enhance the proportion of the detectable proteome.

Protein extraction and preparation

While much attention is paid to developing new technologies for high-resolution protein separation and rapid, automated protein identification, such as state-of-the-art mass spectrometers, the most critical step in any proteomics study is protein extraction and sample preparation. An ideal extraction protocol would reproducibly capture and solubilize the full complement of proteins in a given sample, whilst minimizing post-extraction artifacts and non-proteinaceous contaminants. However, given the diversity of polypeptide molecular size, charge, hydrophobicity, post-translational modification, complexation and cellular distribution, no single protein extraction protocol or solvent system can capture an entire proteome. Consequently, a range of different extraction protocols, involving many permutations of physical treatments, solvents and buffers, have been reported in the literature (Rabilloud, 1996). While these will not be listed here, Figure 2 outlines some common steps or decision-points. The appropriate protocol will largely be guided by the downstream analysis that is to be performed,
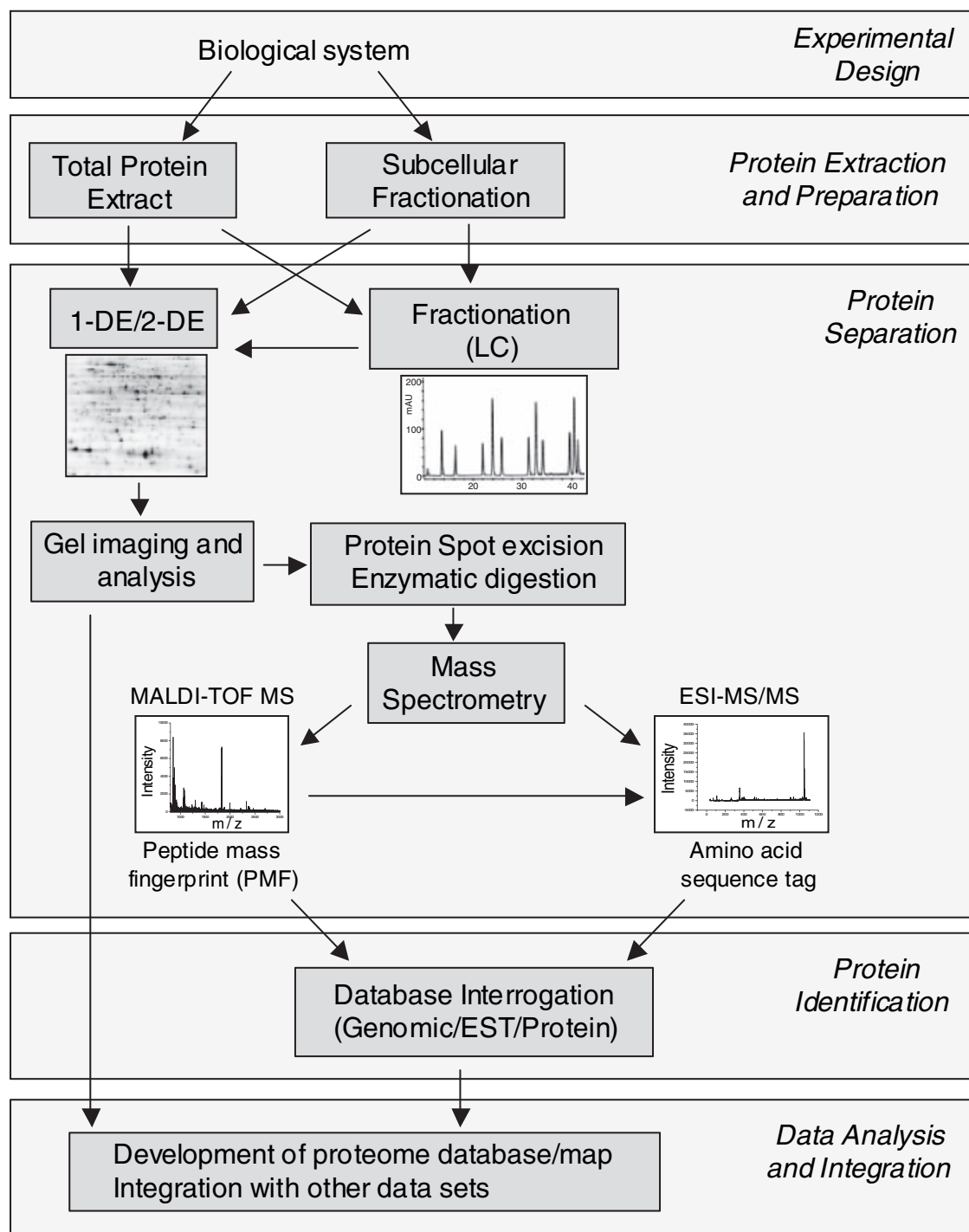
Figure 1. Overview of common steps involved in proteomic analysis. These typically include protein separation by one- or two-dimensional electrophoresis (1-DE or 2-DE, respectively) or liquid chromatography (LC), followed by protein identification using spectra generated by matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) or electrospray ionization tandem mass spectrometry (ESI MS/MS).

although in some instances throughput and speed of sample manipulation, or available facilities, may represent important considerations. For example, in some cases the objective is to identify native protein complexes (e.g. Werhahn and

Braun, 2002), or to combine the extraction with a subsequent assay that requires the proteins to be in their native condition. In this situation, extraction protocols are typically performed with mild chemical agents under non-denaturing
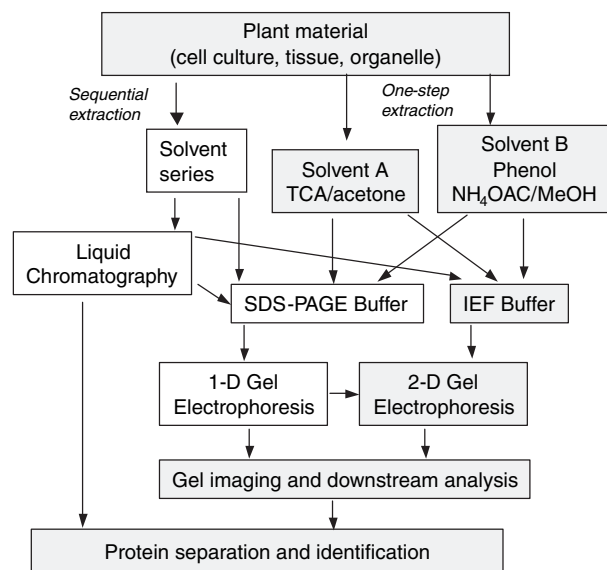
Figure 2. Schematic outline of protein extraction methods. While a wide range of protocols is used, depending on the tissues or experimental design, the shaded boxes represent the most commonly used pathway to obtain 'total protein' extracts prior to comparative proteomic analysis.

conditions (Manabe, 2000) and they can be coupled with downstream blue native polyacrylamide gel electrophoresis; a technique that has been used to fractionate native protein complexes from a range of organisms, including plants (Camacho-Carvajal *et al.*, 2004; Eubel *et al.*, 2003; Giege *et al.*, 2003; Kuchler *et al.*, 2002). In other cases, specialized protocols have been developed to extract specific protein subsets, such as membrane proteins (Everberg *et al.*, 2004; Ferro *et al.*, 2000; Molloy *et al.*, 1998; Santoni *et al.*, 2000a,b), secreted or cell wall-associated proteins (Chivasa *et al.*, 2002; Okushima *et al.*, 2000) or glycosylphosphatidylinosi-tol-anchored proteins (Borner *et al.*, 2003; Sherrier *et al.*, 1999). In addition, sequential extraction of tissues with a series of different solvents (e.g. Maltman *et al.*, 2002; Robertson *et al.*, 1997) is an effective means to subdivide distinct protein populations, thereby decreasing protein complexity, enhancing the detection of low abundance proteins and increasing the overall detectable proportion of the proteome. The same benefits are obviously derived from targeting pure or highly enriched subproteomes of organelles or subcellular compartments (Jung *et al.*, 2000).

However, while the use of a multi-step extraction procedure has a number of benefits, one caveat is that the number of fractions that require subsequent analysis may escalate to an impractical degree once sample and experimental replicates are considered. A second potential problem is one of reproducibility. The physicochemical diversity of proteins in a cellular extract means that protein extracts are inherently unstable. This instability is manifested as both chemical modification, such as proteolytic degradation, and

differential precipitation and loss from the sample through non-specific binding to surfaces such as pipette tips and tubes. Essentially almost any change in the physical or chemical environment of a protein has the potential to alter its stability and solubility: changes in factors such as solvent pH, ionic strength, temperature and intermolecular interactions may all lead to selective losses of specific protein species from a complex protein mixture. Therefore, the greater the number of experimental manipulations from the moment of extraction to the point of protein separation and detection, the correspondingly greater the chance of losing protein subsets. This includes such commonly used procedures as precipitation, dialysis, desalting, ultrafiltration, liquid chromatography and storage at low temperatures. Consequently, while sequential extraction, or the use of specific solvents and procedures to target subsets of the proteome, are vital for some types of analysis, most comparative proteomics strategies aim to capture the most comprehensive possible spectrum of proteins. This typically involves rapid 'one-step' chemical extraction procedures with a stringent solvent cocktail that is capable of disrupting protein aggregates and denaturing the constituent proteins in the minimum number of steps.

Probably the most commonly reported protocol of this type involves protein precipitation with trichloroacetic acid (TCA) and acetone (Figure 2), which increases the protein concentration and helps remove contaminants (Santoni *et al.*, 1994). However, each extraction strategy should also take into consideration the nature of the sample tissue. Plant materials are typically more problematic for proteomic analysis than tissues from other organisms, as in addition to having relatively low protein concentrations, plant cells are often rich in proteases and compounds that severely interfere with downstream protein stability, separation and analysis. These include cell wall and storage polysaccharides, lipids, phenolic compounds and a broad array of secondary metabolites (Gegenheimer, 1990; Granier, 1988; Tsugita and Kamo, 1999). Indeed, the prevalence of these compounds possibly represents the most significant problem associated with plant proteome analysis. The TCA/acetone procedure, while extremely effective for some plant tissues, and particularly so for young growing vegetative tissues, can sometimes result in the co-extraction of polymeric contaminants. This is a particular problem with more mature tissues and those that have high levels of soluble cell wall polysaccharides and polyphenols (Saravanan and Rose, 2004). An alternative protocol (Figure 2) involves the solubilization of proteins in phenol, with or without SDS, and subsequently precipitation with methanol and ammonium acetate (Hurkman and Tanaka, 1986; Meyer *et al.*, 1988). This method can effectively generate high quality protein extracts with minimal apparent contamination, even from resistant tissues such as wood (Mijnsbrugge *et al.*, 2000) and olive leaves, which contain large amounts of polyphenols (Wang

et al., 2003). Another report indicated that the phenol-based method minimizes proteolysis during extraction (Schuster and Davies, 1983). These two protocols have been used separately in proteomic studies of a number of different plant species and tissues but little has been reported to date in terms of a direct comparison between the two. However, a recent study revealed that these approaches result in substantial differences in the spectrum of proteins that are extracted (Saravanan and Rose, 2004). An example is shown in Figure 3, which shows contrasting two-dimensional (2-D) gel analyses of proteins from tomato roots that were extracted using two variants on the TCA–acetone protocol and the phenol-based method. Certain regions of the gels are boxed to highlight the substantial difference in the protein spot patterns between the extracts. All three protocols performed well with some tissues but there were clear qualitative differences in all tissues examined, and while no obvious pattern emerged that might explain the basis for the differential extraction, it was noted that the phenol-based method showed enhanced extraction of glycoproteins (Saravanan and Rose, 2004). These results suggest that the phenol- and TCA/acetone-based methods are complementary, generally robust across plant tissues and combine ease of use with effective removal of contaminating material. Although the phenol extraction method is sometimes described as more time consuming, the sample preparation stage is certainly not the limiting step of proteomic analysis and the benefits of visualizing a substantially greater number of proteins far outweigh the small additional time investment in sample preparation. The addition of sequential solubilization steps with these protocols has also been shown to increase the number of detectable proteins (Jacobs *et al.*, 2001) and it is also likely that other protocols will be developed that will further enhance or complement these approaches.

For comparative proteome profiling to be effective, it has to be assumed that the same protein species exhibit the same behavior during extraction and subsequent analysis. For example, if fruit ripening-related proteins are to be identified and characterized using comparative proteomics, a specific polypeptide that is expressed in a pre-ripe fruit must be as readily extractable and quantitatively detectable as that same polypeptide species in a ripe fruit, otherwise differences in proteins abundance are impossible to discern. To address this concern, our group has recently been using a sequential extraction protocol to isolate proteins that show differing degrees of affinity to the cell wall pellets from homogenized tomato fruit at different developmental stages. Subsequent Western analyses, using a range of antibodies to proteins that are localized in different subcellular compartments indicate that many intracellular proteins from the cytosol and various organelles bind with extremely high affinity to the polysaccharide-rich cell wall residue during extraction (Saravanan and Rose, Cornell University,
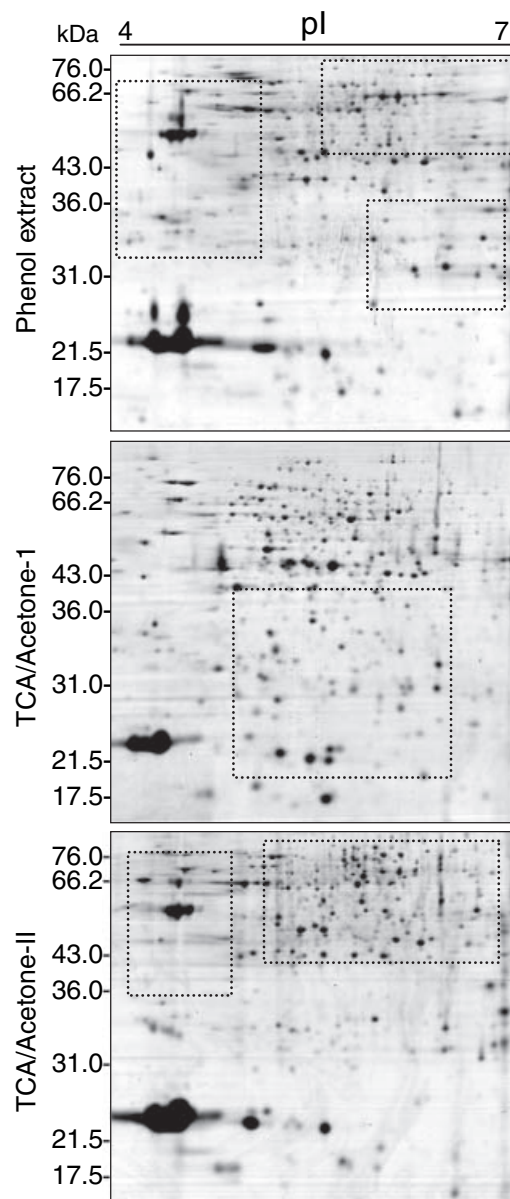

Figure 3. Comparative 2-DE analysis of tomato root proteins using three extraction techniques. Proteins were extracted using two variants of a protocol using TCA/acetone (TCA/Acetone I and TCA/Acetone II) or using a phenol-based buffer (described in Saravanan and Rose, 2004). The gel analyses were performed using pH 4–7 non-linear IPG strips (17 cm) in the first dimension and 12% SDS gels in the second dimension. The gels were stained with colloidal Coomassie blue. The boxed areas highlight areas of the gel that show substantial variability in the spot patterns generated using the three protocols.

Ithaca, NY, USA, unpublished data). Of even more concern is the observation that this artifactual loss appears to vary substantially for a particular protein from tissues at different developmental stages. This likely reflects the substantial differences in wall structure and composition that occur during fruit development and ripening. For example, large

polyanionic pectic polysaccharides from the primary wall and middle lamella undergo substantial modification during fruit ripening (Rose *et al.*, 2003). It is conceivable that such changes contribute to the differential protein binding capacity of the wall residues in homogenized tissues although, interestingly, even the use of high ionic strength extraction buffers does not fully resolve the problem. These results have potentially important implications for both comparative proteomic analysis and also when correlating dynamic changes in protein levels with other data, such as microarray analyses. The concern is clearly that plant biologists may in some cases be evaluating the 'extractome' rather than the proteome. It is not apparent yet whether this phenomenon occurs with all plant tissues, nor whether it is specific to plants, given their unique wall structure, but further studies are in progress. However, it seems that this represents a problem that is unique to plant proteome analysis and one that is typically under-appreciated.

2-D gel electrophoresis: the good, the bad and the ugly

Following extraction, the next objective is to obtain as much qualitative and quantitative information about the constituent protein population as possible. As is indicated in Figure 2, various approaches are possible, including 1-D SDS-PAGE analysis. However, for quantitative comparative proteomic analysis, there is still no widely available technology that surpasses 2-D gel electrophoresis (2-DE), which has been the method of choice for separation of complex protein mixtures for several decades. Several recent reviews have provided a detailed overview of 2-DE and associated sample preparation (e.g. Görg *et al.*, 2000; Lilley *et al.*, 2002; Ong and Pandey, 2001), so this will not be reiterated. However, briefly, 2-DE involves separating proteins in the first dimension based on their charge and in the second dimension based on their molecular mass. The recent increased popularity of 2-DE has been mainly the result of significant improvements in resolution and reproducibility, as summarized in an excellent review by Rabilloud (2002). In particular, the availability of both broad pI range (typically pI 3–10) and narrow range (e.g. 1 pI unit) immobilized pH gradient (IPG) separations in the first dimension has substantially increased the resolution and consequently the number of distinct proteins spots that can be detected (Görg *et al.*, 2000). While IPG-based 2-DE is substantially easier and more reproducible than the original first dimension 'tube gel' format (see Rabilloud, 2002), a degree of technical proficiency is still required and it should be stressed that high quality protein samples are essential as contaminants can dramatically interfere with separation. Websites with useful protocols and troubleshooting guides for 2-DE analysis are listed in Table 2.

2-DE is by no means the only platform for protein separation and there are inherent technical limitations, such as the limited ability to fractionate specific classes of proteins including hydrophobic proteins and glycoproteins, or to visualize low abundance proteins (Harry *et al.*, 2000; Rabilloud, 2002). For example, it is well established that hydrophobic membrane proteins do not fractionate well using 2-DE (e.g. Santoni *et al.*, 1999, 2000a) and are rarely present in lists of 2-DE-derived proteins from any species. While this is an important topic, given that transmembrane proteins have been estimated to comprise 30% of total cellular proteins (Santoni *et al.*, 2000a), it is not one that is specific to plants and so will not be discussed in detail here. 2-DE is also notoriously difficult to automate (Lopez, 2000; Quadroni and James, 1999), which limits throughput and results in greater experimental variability through manual

Resource	URL
Plant proteomics databases	http://www.ncbi.nlm.nih.gov
	http://www.mips.biochem.mpg.de/
	http://www.pir.georgetown.edu/
	http://www.pir.georgetown.edu/pirwww/pirnref.shtml
	http://www.prf.or.jp/en/
	http://www.au.expasy.org/
Subcellular or organelle databases	http://www.mitoz.bcs.uwa.edu.au/apmdb/APMDB_Database.php
	http://www.cbsusrv01.tc.cornell.edu/users/ppdb/
	http://www.megasun.bch.umontreal.ca/gobase/gobase.html
	http://www.sphinx.rug.ac.be:8080/ppmdb/index.html
	http://www.genomics.msu.edu/plant_specific/index.html
	http://www.gartenbau.uni-hannover.de/genetik/AMPP
	http://www.aramemnon.botanik.uni-koeln.de/
	http://www.lecb.ncifcrf.gov/2dwgDB/
2-DE database	http://www.expasy.ch/ch2d/2d-index.html
	http://www.aestivum.moulon.inra.fr/imgd/
	http://www.pierroton.inra.fr/genetics/2D/index.html
	http://www.swissproteomicsociety.org/links.html
	http://www.uia.ac.be/tby2/

Table 2 Online plant proteomics-related resources

intervention. Alternative 'gel-less' approaches, such as multi-dimensional protein identification technology (MUD-PIT), have already been used effectively to catalog many polypeptides in total protein mixtures in several organisms, including rice (Koller *et al.*, 2002; Whitelegge, 2002). However, while MUDPIT is an excellent means to generate an exhaustive catalog of proteins present in a particular protein sample, it does not yield reproducible quantitative information. The isotope coded affinity tag (ICAT) technique has recently been developed to improve quantitative comparisons in the absence of 2-DE (Gygi *et al.*, 1999b). However, the reproducibility and the number of replicates required for determining statistical significance have yet to be completely resolved (Rabilloud, 2002). While ICAT is being adopted by a rapidly growing number of laboratories, no major ICAT analysis of plant proteins has been reported and as a new technology still under evaluation, it will not be discussed in detail here. In the context of comparative proteomics then, where the goal is to identify quantitative and qualitative differences between protein samples, a 2-DE approach is currently the method of choice, as it generates data in a form that allows far easier visual evaluation and allows quantitative comparisons (Rabilloud, 2002).

Several relatively inexpensive commercial 2-DE systems (in the region of $10 000 for a complete system) are available (e.g. from Amersham Biosciences, Bio-Rad and Invitrogen: note that throughout this review no particular product or company is endorsed and such lists are not intended to be complete, but are merely to provide examples). The first dimension isoelectric focusing (IEF) apparatus can typically accommodate IPG focusing strips of various sizes, from approximately 7 cm to larger than 20 cm. Corresponding sizes of second dimension SDS-PAGE systems are also generally available from the same vendors that can be used with either manually cast or pre-cast commercially available SDS-PAGE gels. Either option is effective, although for final sample preparation or critically important separations, where a high-quality, uniform gel matrix is particularly important, commercial pre-cast gels are often a preferred and convenient option. In practice, while the smaller gels can provide a useful preliminary qualitative screen of protein extracts, larger gels of at least 17 cm in the first dimension are more useful for comparative studies as they have a greater loading capacity, resulting in a greater number of spots and give substantially better spot separation. While some publications have reported up to 10 000 protein spots in a single gel (Hatzimanikatis *et al.*, 1999b; Klose, 1999), this is unusual and, depending on loading and sample quality, single 2-D gel images generally contain approximately 1000–2000 distinguishable protein spots. Clearly this does not even closely approach the expected proteome of a tissue and it has been estimated that single gel-based analyses allow identification of approximately 5% of expressed cellular proteins (Heazlewood and Millar, 2003). Obviously, this number is in stark contrast to the percentage of expressed genes that might be detectable using a transcript profiling approach.

However, the proportion of the accessible proteome can be increased through the use of multiple overlapping narrow-range IPG separations (Hoving *et al.*, 2000; Wildgruber *et al.*, 2000) or by incorporating prefractionation steps. This can take the form of biological prefractionation, targeting organelles or subcellular fractions, or biochemical prefractionation such as with multi-compartment IEF systems (Hamdan and Righetti, 2003) or column chromatography-based separations (Figure 2). While such approaches can certainly increase the number of detectable proteins when performing comparative studies, a couple of points should be taken into account when considering this higher resolution strategy. First, at the practical level, as the number of pre-fractions or narrow pI range separations increases, the consequent workload rapidly escalates, particularly when taking into account the replicate gels that must be run for every separation to account for experimental variability, as well as the replicates to evaluate biological variability. For example, if two tissue samples are to be compared using only three different pI ranges (e.g. 3–10, 4.5–5.5 and 5.5–6.7) and if five replicate gels are run per IEF separation to improve the quantification of protein abundance, this would require 30 individual 2-D gels. As most studies would evaluate biological variability, replicate sets of tissues are typically run, corresponding to multiples of 30 gels. The cost and time investment in such an exercise can therefore be substantial (Lopez, 2000).

A second potential problem with the higher resolution pre-fractionation approach is that, as previously mentioned, every additional sample manipulation step runs the risk of causing non-specific loss of certain proteins. For example, column chromatography of complex protein samples is not perfectly reproducible and a specific protein may fractionate differently between runs and between samples. When performing comparative proteomic analysis, the greater the number of extraction and preparation steps, the greater the chance of missing or erroneously detecting a differentially expressed protein.

Alternative gel-less protein fractionation systems are starting to appear on the market (e.g. http://www.beckman-coulter.com/products/instrument/protein/proteomelab_pf2d_dcr.asp) that are based entirely on chromatographic separation and that claim particularly high reproducibility, thus allowing comparative studies to be made. These may prove to be useful complements to existing approaches, but 2-DE remains the most effective proven approach for comparative proteomic analysis (Rabilloud, 2002).

Gel staining, imaging and analysis

Following gel electrophoresis, the next steps are protein staining and image analysis in order to quantify each protein
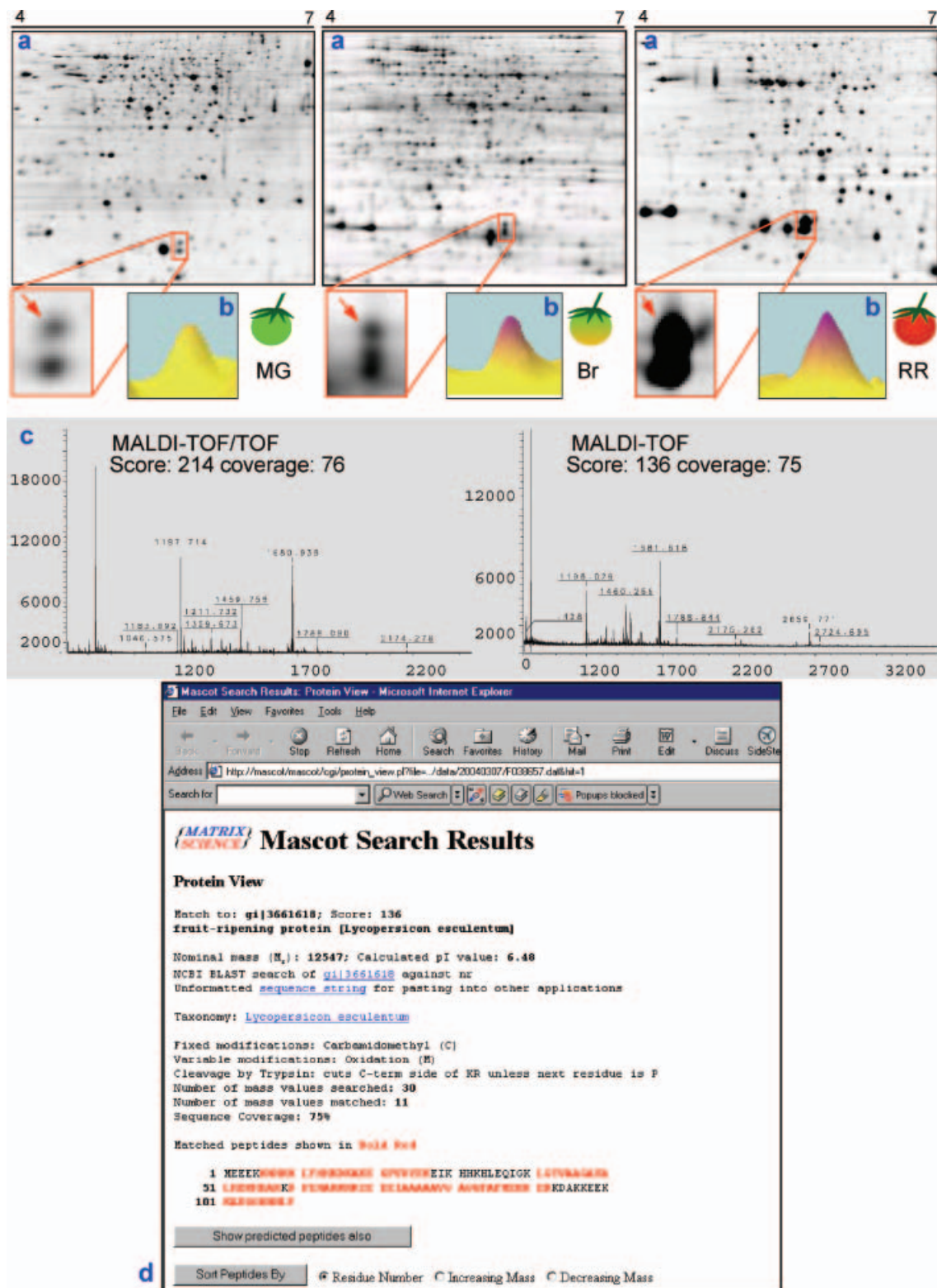
and to allow qualitative comparison of samples. Here again, technical hurdles restrict the quantification of all proteins in any one sample, as the range of protein concentrations can often span more than seven orders of magnitude, which is beyond the accurate measure of any staining or quantification system. For many years, two basic options were typically followed for protein staining (Patton, 2000a). Colloidal Coomassie staining is relatively easy, cost-effective and compatible with subsequent protein identification by mass spectrometry (MS), but it is only moderately sensitive, with a limit of approximately 10 ng protein. The other alternative has been silver-staining, which is more sensitive, detecting as little as 0.5 ng protein (Heazlewood and Millar, 2003) but not particularly quantitative and less suitable for MS identification (Lopez, 2000; Patton, 2000a). More recently, a number of sensitive fluorescent stains, such as SYPRO Ruby and SYPRO Orange have been developed that combine the advantages of the other stains; a similar sensitivity to silver stain, but the ease of use and excellent MS compatibility (Lauber *et al.*, 2001; Yan *et al.*, 2000) of Coomassie. However, in some cases the high costs of dyes such as SYPRO Ruby may be prohibitive, particularly with large projects involving many gels.

Gel imaging is usually performed with a laser scanner or CCD-based system (reviewed in Miller *et al.*, 2001; Patton, 2000a,b). Laser scanners typically operate more slowly, but achieve better spatial resolution than standard CCD cameras. In some cases with Coomassie or silver stains, cheaper flatbed scanners allow rapid gel imaging and documentation, but the quantification is less accurate than with a laser scanner. Cooled CCD cameras have improved signal to noise ratios and also typically claim linear quantification over three to four orders of magnitude (Patton, 2000b).

Following gel imaging, the next step in the workflow (Figure 1) is image editing and comparative analysis. This represents a major bottleneck in any comparative proteomics project and although several commercially available software packages are available to assist with image acquisition, spot editing, quantification, annotation, comparisons and generation of web-formatted data sets, this step requires substantial manual intervention and time investment. Commonly used software packages include Melanie 4 (http://www.expasy.org/melanie), which is now incorporated into ImageMaster 2D Platinum version 5 (http://www.amershambiosciences.com), PDQuest (http://www.

proteomeworks.bio-rad.com/html/pdquest.html), Phoretix and Progenesis (http://www.nonlinear.com/products/2d) and Z3 and Z4000 (http://www.2dgels.com). Such packages vary in cost by an order of magnitude (from approximately $10 000 to more than $100 000), depending on the degree of automation, analytical speed and 'bells and whistles', but regardless of whether an entry-level or upper-level package is used, they require considerable time for familiarization with all the strengths or foibles. This technology is certainly not at a point where it can be said to be highly automated.

To illustrate the kind of data that are generated using the approach described above, Figure 4 shows a 2-DE-based comparative analysis of tomato fruit pericarp during ripening. Three stages of fruit development were selected: mature green (MG; the stage immediately prior to the onset of ripening), breaker (Br; the stage that marks the early onset of ripening) and red ripe (RR; a late ripening stage). Proteins were extracted using a variant of the phenol-based method, which is particularly effective with tissues such as fruit pericarp that are rich in charged polysaccharides (Saravanan and Rose, 2004). The panels labeled 'a' show 2-D gels with a pI range of 4–7, stained with colloidal Coomassie blue. Progenesis (Non-Linear Dynamics) software was used to identify ripening regulated proteins and examples of two such spots are shown in the magnified panels. Three-dimensional representations of the lower spots, generated using Progenesis, are shown in the panels labeled 'b'. Descriptions of the other panels are given in the MS and database analysis sections below.

To summarize, 2-DE can provide an excellent means of comparing the expression of hundreds proteins between samples and of revealing which show quantitatively substantial differences in expression. The quality of the protein extract is of supreme importance and most unsuccessful attempts to visualize large numbers of well-resolved spots with minimal streaking or smearing are the result of problems with extraction and subsequent preparation. Careful laboratory practices and a degree of experience and technical expertise can result in good reproducibility and accurate spot matching (Chloe and Lee, 2003), without which it is impossible to derive statistically useful information or compare the expression levels of a specific protein in two or more protein populations. This objective is central to most proteome studies and yet remains the most challenging. Inter-gel variation and many other subtle variables

Figure 4. Steps involved in comparative proteomic analysis of tomato fruit ripening.
(a) 2-D gel images of proteins extracted from tomato fruit pericarp tissue at the mature green (MG), breaker (Br) and red ripe (RR) developmental stages, with expanded panels and an arrow highlighting an example of a ripening upregulated protein spot.
(b) Three-dimensional representation of the relative expression levels of the same protein spot generated using Progenesis (Non-Linear Dynamics) image analysis software.
(c) A comparison of mass spectra derived from the highlighted protein spot following trypsin digestion and analysis by MALDI-TOF or MALDI-TOF-TOF. The probability scores and percentage sequence coverage by the resultant peptides are shown, following databases searched using the Mascot search engine.
(d) Screen capture of Mascot search results, revealing the predicted identification of a protein corresponding to a previously identified tomato ripening related gene.

substantially complicate the superimposition of multiple gel images. Commercial image analysis software rarely achieves a high degree of accuracy and comparative studies have to be performed and verified manually. Moreover, this procedure is inherently laborious and prone to errors.

An important breakthrough was made in addressing this rate-limiting step (Unlu *et al.*, 1997) through the development of fluorescent 2-D difference in-gel electrophoresis (DIGE). This technique involves the covalent labeling of two different protein extracts (e.g. from pre-ripe fruit and red ripe fruit, as shown in Figure 5) with one of two fluorescent cyanine (Cy) dyes (Cy3 and Cy5), similar to those used for differential cDNA microarray analysis, but optimized for 2-DE. The two labeled protein samples are then mixed, separated on the same 2-D gel and scanned on a variable wavelength laser-based imaging system. Cy3 and Cy5
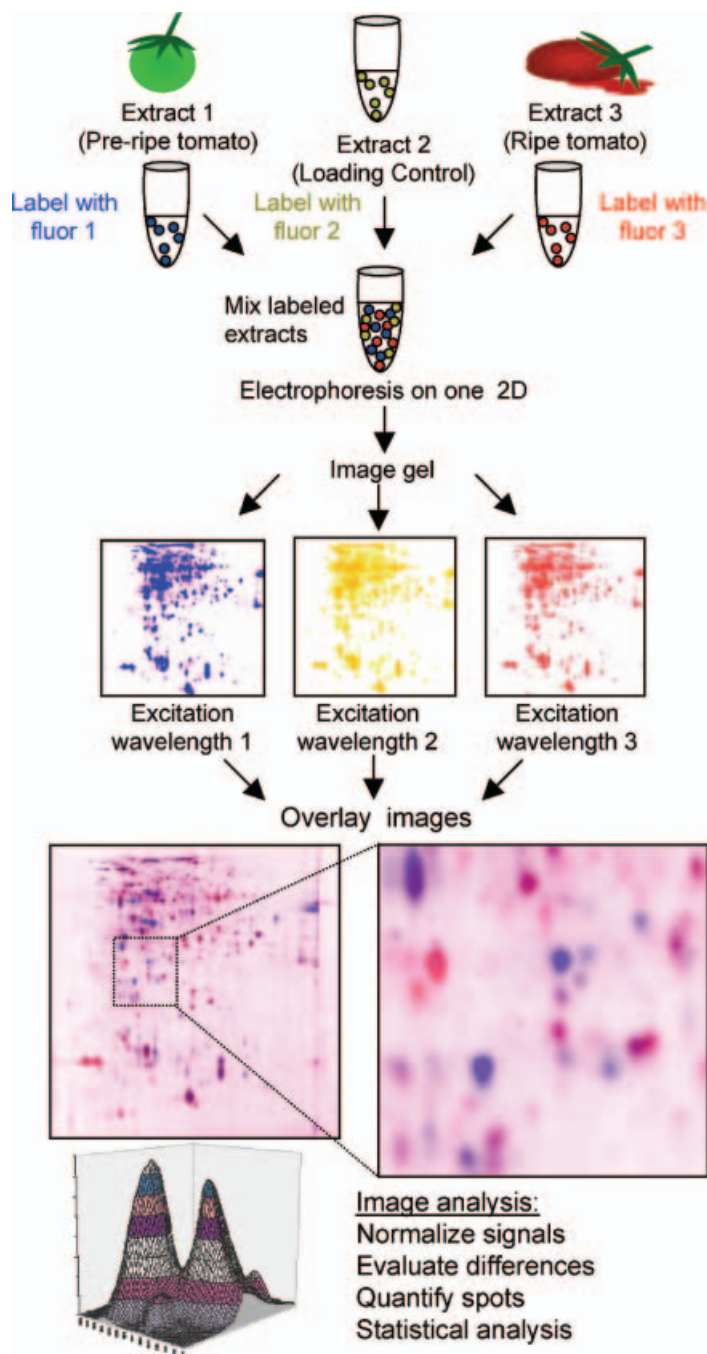

Figure 5. Outline of the difference in-gel electrophoresis (DIGE) approach for comparative proteomics. Samples from three sources (in this case a pre-ripe and a ripe tomato fruit and an internal loading control) are covalently labeled with one of the three Cy dyes (Cy2, Cy3 and Cy5); the samples are pooled, and then separated on a single 2-D gel. Imaging of the gels at different wavelengths, corresponding to the emission spectra of the three dyes, allows a quantitative and qualitative comparison of the protein populations in the original samples and the differential images may be readily analyzed to determine statistical differences. An example of DIGE gel of proteins from pre-ripe and ripe tomato fruit is shown with a portion of the gel amplified to highlight clear blue or red spots that represent differentially expressed proteins.

exhibit distinct excitation and emission spectra and so it is possible to rapidly quantify and distinguish between proteins that were present in either of the original two extracts. As the two protein samples are separated on the same gel, any protein that exists in both populations will migrate to the same location on the 2-D gel, dramatically facilitating comparisons of protein expression in the two original samples. A third fluorescent dye (Cy2) can be used with a third protein extract in the same gel to provide an internal standard for sample normalization, which allows a much more accurate statistical analysis of protein expression across multiple DIGE gels. The issue of the number of gel replicates that are theoretically necessary to draw statistically significant conclusions about differential protein expression is a thorny one and there does not appear to be a consensus opinion. However, DIGE certainly reduces the number. This was recently highlighted in a DIGE-based study where a majority of the identified differentially expressed proteins would have been overlooked in the absence of the internal standard (Friedman *et al.*, 2004). As with microarray analysis, the best approach to validate the expression information is probably to supplement the global expression profiling with additional independent techniques. For example, for gel-based protein identification and quantification this might include Western blot analysis.

Reports to date describing the use of DIGE (Alban *et al.*, 2003; Gharbi *et al.*, 2002; Tonge *et al.*, 2001; Zhou *et al.*, 2002) have demonstrated the potential of this technique and confirm that the Cy dyes have a similar linear dynamic range to the most sensitive and quantitative existing fluorescent stain, SYPRO Ruby. The optimal procedure for gel analysis has been suggested to be initial Cy3/Cy5/Cy2 labeling to obtain images for statistical analysis, followed by post-staining with SYPRO Ruby to visualize the maximum number of spots on one or more preparative gels for spot picking (Gharbi *et al.*, 2002). To date we are aware of only one published report that has described the use of DIGE to study plant proteins (Kubis *et al.*, 2003), although our group is currently using DIGE in a comparative proteomics study of tomato fruit development and ripening.

Protein identification

Continuing down the comparative proteomics pipeline (Figure 1), once a set of differentially expressed spots has been identified from a series of 2-D gels, the next step is typically to identify the cognate proteins and genes. While this can be achieved in a number of ways (Gevaert and Vandekerckhove, 2000), MS is now firmly entrenched as the method of choice for both protein identification and characterization of post-translational modifications. Over the last decade or so, mass spectrometers have become increasingly attractive analytical instruments for biologists, due in part to new ionization methods and major improvements in

mass accuracy, resolution, sensitivity and ease of use, which have extended the applicability of MS to characterize large intact macromolecules such as proteins. The rapid emergence of genomics and proteomics as highly funded and potentially profitable disciplines in the life sciences has provided additional impetus to the development of user-friendly, highly automated mass spectrometers. A diverse range of such instruments is now available that are specifically designed to serve the needs of proteomics researchers, in that emphasis is placed not only on technical performance, such as mass accuracy and sensitivity, but also on automation. Critically, newer spectrometers are packaged with software that facilitate protein identification and structural analysis and that provide a bridge between mass spectra and public sequence databases. Given the frequent specialist reviews describing MS instrumentation, software and techniques (e.g. Aebersold and Mann, 2003; Ferguson and Smith, 2003; Lin *et al.*, 2003; Mann *et al.*, 2001; Standing, 2003) only a brief overview is provided here.

The first step toward protein identification is typically excision of 2-D gel plugs containing the protein spots of interest, in-gel digestion with a site-specific protease (commonly trypsin), and finally MS analysis of the resultant eluted peptides (Figure 1). Two MS platforms in particular represent powerful tools for proteomic studies. The first, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS is typically used to measure the masses of the peptides derived from the trypsinized parent protein spot, generating a 'peptide mass fingerprint' (PMF). Several software packages are then available that can compare the peptide mass list with a predicted 'theoretical' list of tryptic peptide fragments for every protein in the public databases, together with equivalent translated genomic and EST databases. In this case the protein is identified based on the *in silico* match of experimentally determined versus predicted peptide masses, together in some cases with the apparent and predicted pI and molecular mass from the 2-D gels, rather than actual amino acid sequence. Although each PMF is usually a viable means of assigning identity to a specific protein, as a result of the variability in amino acid sequences and the relative distribution of protease cleavage sites between proteins (Godovac-Zimmermann and Brown, 2001) members of protein families with a high degree of sequence similarity can also result in effectively indistinguishable PMFs. This problem is exacerbated by the fact that it is unusual for the full complement of peptides for any given protein to be ionized and detected experimentally by MALDI-TOF.

A common alternative to the PMF approach is *de novo* sequencing by electrospray ionization tandem mass spectrometry (ESI) MS/MS, which yields amino acid sequences of selected tryptic peptides (Figure 1). The first step of tandem MS involves ionization of a sample and separation based upon the mass-to-charge ratio (*m/z*) of the primary ions. An

ion with a specific *m/z* value is then selected, fragmented, and the fragment ions detected after passing through the second mass spectrometer. This process produces a series of fragment ions that can differ by single amino acids, allowing a portion of the peptide sequence, termed an 'amino acid sequence tag', to be determined and used for database searching. When a peptide has been identified in the database, the theoretical fragmentation pattern can be predicted and compared with the observed MS/MS spectrum for assignment of other peaks that can validate the identification. This procedure can be repeated for every fragmented peptide in the sample, leading to additional verification or identification of other proteins in the sample. A number of different types of mass spectrometers can perform MS/MS (reviewed by Aebersold and Goodlett, 2001; Handley, 2001; Roberts, 2002) including triple quadrupole (Mann and Wilm, 1995; Wilm and Mann, 1996) and MALDI-TOF/TOF instruments (Bienvenut *et al.*, 2002). Panel 'c' in Figure 4 shows an example of both MALDI-TOF and MALDI-TOF/TOF mass spectra derived from the tomato fruit ripening upregulated protein shown in the 'b' panels. The spectra were used with a software package (Mascot, Matrix Science) to identify the protein. Screen-captures of the results are shown in panels 'd' and 'e', which indicate that 11 peptides were matched between theoretical and observed data sets, corresponding to 75% coverage of a protein encoded by a previously identified ripening-related tomato gene. A number of other search engines that can perform similar analyses of MS spectra in conjunction with sequence databases are listed in Table 1.

The pros and cons of each approach have been described at length in the MS reviews cited above and further details are provided in a recent plant proteomics review (Heazlewood and Millar, 2003). To summarize, the MALDI-TOF/PMF approach is very rapid, requires relatively little user expertise, can be automated and is tolerant of contaminants. On the contrary, the data are generally more ambiguous and essentially rely on the availability of a genomic sequence, or at least a substantial EST collection, for the species under study. In contrast, MS/MS analysis is technically more challenging, requires specialist training, has a slower throughput and is far more expensive. However, it is generally considered to provide a more conclusive 'answer' in terms of definitively identifying a protein, as well as the obvious value of allowing the study of post-translational modifications and other structural features.

The high cost of mass spectrometers obviously places them well beyond a routine laboratory purchase, but many universities are now establishing MS and proteomics facilities, much as automated DNA sequencing facilities are now commonplace and some examples are listed in Table 1. Such centers typically have both MALDI-TOF and MS/MS instrumentation, together with other proteomics-related equipment, and operate as either a 'drop-off' service, or

can provide training for customer-operated analysis with an hourly fee. These facilities often offer protein sequencing and identification to the general research community beyond the immediate campus, as do a number of companies that provide a range of proteome analysis services, from gel-separation and analysis to protein identification and structural characterization. Such services are not cheap but, as with DNA sequencing, the cost of mass spectrometers and MS analysis is decreasing as instrumentation, automation and ease of use improves.

Comparative proteomics of 'non-model' plant species

Proteomics is built on a foundation of DNA sequence and the recent completion of the first two plant genome sequences, *Arabidopsis thaliana* (The *Arabidopsis* Genome Initiative, 2000) and rice (Goff *et al.*, 2002; Yu *et al.*, 2002), in addition to large-scale plant EST sequencing initiatives (http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html), are already catalyzing substantial developments in the field of plant proteomics. This rapidly growing repository of plant DNA sequence information provides a means to associate a protein sequence with the cognate nucleotide sequence and so the increasing diversity of plant genome and EST sequencing projects will enable parallel proteomics analyses. Thus, the first published attempts to characterize plant proteomes, which were developed in species such as *Arabidopsis* and rice at a time when the genome sequence was not available (Komatsu *et al.*, 1999; Tsugita *et al.*, 1996), typically resulted in the identification of relatively few proteins. In contrast, recent proteomic studies of plant tissues such as maize leaves (Porubleva *et al.*, 2001), *Arabidopsis* seeds (Gallardo *et al.*, 2001), rice embryos (Woo *et al.*, 2002) and *Medicago truncatula* roots (Mathesius *et al.*, 2002) and organ-specific proteomics of *M. truncatula* (Watson *et al.*, 2003) have identified hundreds of proteins. Moreover, a recent report described the first substantial shotgun sequencing of plant proteins from whole tissues, where 2500 unique rice proteins were identified (Koller *et al.*, 2002). Clearly the field of proteomics has matured to a point where it has become a highly attractive experimental platform for a broad range of plant biologists. However, a key question is whether the substantial investments that have been made in developing a genomics infrastructure in specific plants species, such as *Arabidopsis* and rice, will facilitate proteomic studies across the entirety of the plant kingdom.

An example of such broad applicability would be the ability to deduce the identity of a protein spot by comparing its location on a 2-D gel with an equivalent 2-D gel reference map of sequenced and identified proteins from another species. To this end, open-source software is available that allows comparisons of 2-D gel images with annotated databases 2-D gels that have been run under the same

standardized conditions (http://www.lecb.ncifcrf.gov/flicker). In theory, the greater the DNA sequence similarity between organisms, the greater the predicted similarity of the 2-D gel spot pattern. A study contrasting proteins from two closely related bacterial species and comparing total protein patterns from three legumes indeed revealed marked similarities in the positions of predicted orthologous proteins on 2-D gels (e.g. Mathesius *et al.*, 2002). However, given the substantial effects that small differences in pI and post-translational modification can have on the migration of a protein in a 2-D gel, and the high possibility of co-migration of unrelated proteins, gel position alone should certainly not be taken as conclusive evidence of identity and supporting data are needed. Another related issue that may also substantially limit the potential of 2-D gel spot pattern matching as a predictor of protein identity is that in many cases the patterns of protein spots from closely related species are remarkably different. This is exemplified in Figure 6, which shows a 2-D gel analysis of proteins from the pericarp of tomato (*Lycopersicon esculentum*) and pepper (*Capsicum annuum*) fruit at the onset of ripening. Despite the fact that these solanaceous species are very closely related, and that the protein extracts came from a specific tissue at an equivalent developmental stage, the overall patterns are substantially different. Few clusters of spots appears to be conserved and of the more than 1000 detectable protein spots per gel, only 30–40% appeared to show a similar gel location, many of which may correspond to completely unrelated proteins. Differences in gene expression between ripening tomato and pepper may contribute in part to the dissimilar protein spot patterns, but sufficient fundamental ripening-related processes are con-

served between tomato and pepper, that a greater similarity might be expected (Saravanan and Rose, unpublished data). Possible explanations, such as differential protein extractability and stability are currently being evaluated, but clearly at present even a highly annotated gel map of tomato fruit proteins would have little direct benefit for identifying orthologous pepper proteins and not surprisingly, equivalent 2-D gel patterns of fruits from more divergent species show even less similarity (Saravanan and Rose, 2004).

As previously mentioned, the ability to identify proteins, whether using amino acid sequence tags or PMF data or a combination of the two, is enhanced by the availability of an appropriate DNA sequence data set (Heazlewood and Millar, 2003). An important question then is whether a substantial DNA sequence repository, whether genomic sequence or high-coverage unigene set derived from a large EST collection, in a given species permits accurate protein identification in a related species. A related question is if this is possible, how closely related do those species have to be? The substantially greater speed and ease of acquiring PMF data by MALDI-TOF MS compared with *de novo* sequencing using an ESI-MS/MS approach means that successful cross-species PMF studies would be particularly desirable. A study that directly addressed this question using PMF analysis of proteins from four plant species (Mathesius *et al.*, 2002) concluded that PMF data are not particularly useful for cross-species protein identification except for highly conserved proteins. Furthermore, the confidence that can be assigned to a database hit using PMF data is strongly dependent on the nature of the DNA database: EST databases typically contain a significant proportion of incomplete sequences which reduces the information content of the search and
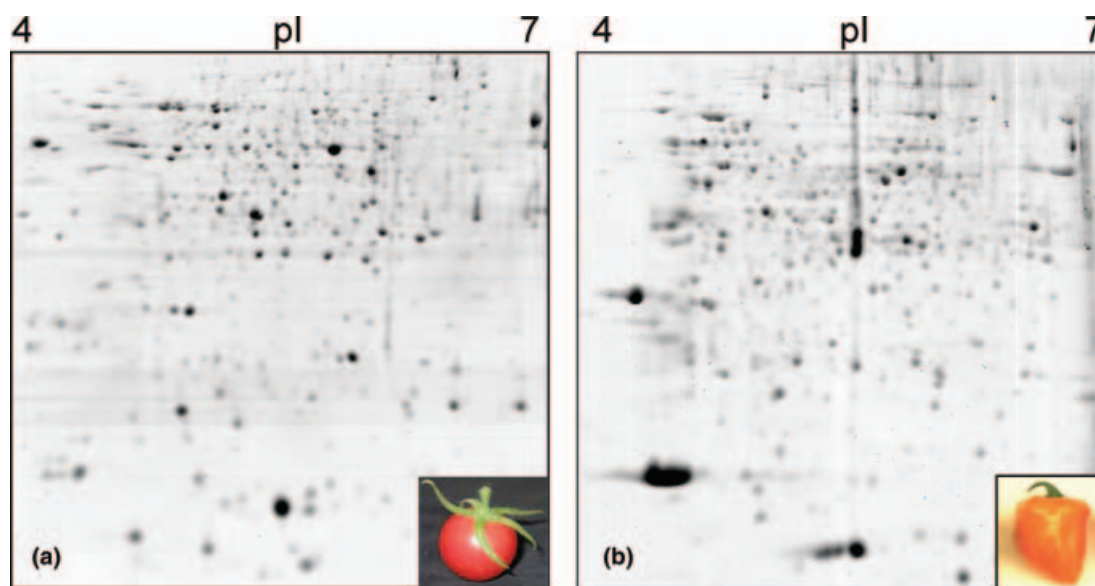

Figure 6. 2-DE analysis of proteins from tomato and pepper fruit. Proteins were extracted from ripe tomato (a) and pepper (b) fruits and subjected to 2-DE analysis using pH 4–7 non-linear IPG strips (17 cm) in the first dimension and 12% SDS gels in the second dimension. The gels were stained colloidal Coomassie blue.

hence the probability score. However, even having a complete genomic sequence available, as is now the case with *Arabidopsis* and rice, generally results in a high confidence identification only approximately half the time with PMF data, while MS/MS-spectra are more likely to result in a positive match. This issue was neatly summarized in recent review (Heazlewood and Millar, 2003). A common strategy to identify proteins is therefore to use MALDI-TOF as a high throughput screen to obtain PMFs and MS/MS, or related *de novo* sequencing MS approaches, as a second screen to reanalyze samples that were not identified in the first MS screen. Recently, a new generation of MALDI-TOF/TOF instruments has been developed to generate high throughput MS/MS data and PMF data from the same sample (Bienvenut *et al.*, 2002; Medzihradszky *et al.*, 2000).

Conclusions

The quality and quantity of data emerging from plant transcript-profiling initiatives (see Alba *et al.*, 2004) has fueled similar expectations for large-scale, high-throughput protein profiling (Kersten *et al.*, 2002). However, while much progress has been made in plant proteomics, it is important to consider the theoretical and practical limitations and to have an appreciation of what can and cannot currently be achieved when evaluating plant protein expression, particularly when considering cost and availability of the specialist instrumentation that is typically required. This review outlines some of the existing challenges and rate limiting steps, which span protein extraction, separation and the critical importance of an extensive affiliated DNA sequence data set and bioinformatics support.

While quantitative analysis of a major portion of a plant proteome is not yet within reach, it is currently possible, at a reasonable cost, to undertake a comparative protein profiling analysis and to assess expression of many hundreds of polypeptides. This narrow window on the proteome may not provide the more comprehensive survey that is afforded by microarray analysis, but the advantages of surveying protein expression still more than justify this approach. In this sense, transcript and protein profiling are currently complementary, rather than equivalent fields and are typically used to answer different biological questions. This disparity will decrease as future technologies increase the breadth of the proteome that can be studied, but it is likely to be some considerable time before plant transcriptome and proteome analyses can be fully and seamlessly integrated.

Acknowledgements

References

Aebersold, R. and Goodlett, D.R. (2001) Mass spectrometry in proteomics. *Chem. Rev.* **101**, 269–295.

Aebersold, R. and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*, **422**, 198–207.

Alba, R., Fei, Z, Payton, P., Liu, Y., Moore, S.L., Debbie, P., Gordon, J.S., Rose, J.K.C., Martin, G., Tanksley, S.D., Bouzayen, M., Jahn, M.M. and Giovannoni, J. (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J.* doi: 10.1111/j.1365-313X.2004.02178.x

Alban, A., David, S.O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S. and Currie, I. (2003) A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics*, **3**, 36–44.

Anderson, N.L. and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*, **19**, 1853–1861.

Anderson, N.L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, **18**, 533–537.

Andon, N.L., Hollingworth, S., Koller, A., Greenland, A.J., Yates, J.R. and Haynes, P.A. (2002) Proteomic characterization of wheat amyloplasts using identification of proteins by tandem mass spectrometry. *Proteomics*, **2**, 1156–1168.

Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant. *Arabidopsis thaliana. Nature*, **408**, 796–815.

Bardel, J., Louwagie, M., Jaquinod, M., Jourdain, A., Luche, S., Rabilloud, T., Macherel, D., Garin, J. and Bourguignon, J. (2002) A survey of the plant mitochondrial proteome in relation to development. *Proteomics*, **2**, 880–898.

Bienvenut, W.V., Deon, C., Pasquarello, C., Campbell, J.M., Sanchez, J.C., Vestal, M.L. and Hochstrasser, D.F. (2002) Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics*, **2**, 868–876.

Borner, G.H.H., Lilley, K.S., Stevens, T.J. and Dupree, P. (2003) Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. *Plant Physiol.* **132**, 568–577.

Camacho-Carvajal, M.M., Wollscheid, B., Aebersold, R., Steimle, V. and Schamel, W.W.A. (2004) Two-dimensional blue native/SDS gel electrophoresis of multi-protein complexes from whole cellular lysates – a proteomics approach. *Mol. Cell. Proteomics*, **3**, 176–182.

Canovas, F.M., Dumas-Gaudot, E., Recorbet, G., Jorrin, J., Mock, H.P. and Rossignol, M. (2004) Plant proteome analysis. *Proteomics*, **4**, 285–298.

Chivasa, S., Ndimba, B.K., Simon, W.J., Robertson, D., Yu, X.L., Knox, J.P., Bolwell, P. and Slabas, A.R. (2002) Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis*, **23**, 1754–1765.

Chloe, L.H. and Lee, K.H. (2003) Quantitative and qualitative measure of intralaboratory and two-dimensional protein gel reproducibility and the effects of sample preparation, sample load, and image analysis. *Electrophoresis*, **24**, 3500–3507.

Cho, S.Y., Park, S.G., Lee, D.H. and Park, B.C. (2004) Protein–protein interaction networks: from interactions to networks. *J. Biochem. Mol. Biol.* **37**, 45–52.

Dreger, M. (2003) Subcellular proteomics. *Mass Spec. Rev.* **22**, 27–56.

Ellis, R.J. (1979) The most abundant protein in the world. *Trends Biochem. Sci.* **4**, 241–244.

Eubel, H., Jansch, L. and Braun, H.P. (2003) New insights into the respiratory chain of plant mitochondria. Supercomplexes and a unique composition of complex II. *Plant Physiol.* **133**, 274–286.

Everberg, H., Sivars, U., Emanuelsson, C., Persson, C., Englund, A.K., Haneskog, L., Lipniunas, P., Jornten-Karlsson, M. and Tjerneld, F. (2004) Protein pre-fractionation in detergent-polymer aqueous two-phase systems for facilitated proteomic studies of membrane proteins. *J. Chromatogr. A*, **1029**, 113–124.

Ferguson, P.L. and Smith, R.D. (2003) Proteome analysis by mass spectrometry. *Ann. Rev. Biophys. Biomol. Struct.* **32**, 399–424.

Ferro, M., Seigneurin-Berny, D., Rolland, N., Chapel, A., Salvi, D., Garin, J. and Joyard, J. (2000) Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. *Electrophoresis*, **21**, 3517–3526.

Friedman, D.B., Hill, S., Keller, J.W., Merchant, N.B., Levy, S.E., Coffey, R.J. and Caprioli, R.M. (2004) Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics*, **4**, 793–811.

Fukao, Y., Hayashi, M. and Nishimura, M. (2002) Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of *Arabidopsis thaliana. Plant Cell Physiol.* **43**, 689–696.

Gallardo, K., Job, C., Groot, S.P.C., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. (2001) Proteomic analysis of *Arabidopsis* seed germination and priming. *Plant Physiol.* **126**, 835–848.

Ge, H., Walhout, A.J.M. and Vidal, M. (2003) Integrating 'omic' information: a bridge between genomics and systems biology. *Trends Genet.* **19**, 551–560.

Gegenheimer, P. (1990) Preparation of extracts from plants. *Meth. Enzymol.* **182**, 174–193.

Gevaert, K. and Vandekerckhove, J. (2000) Protein identification methods in proteomics. *Electrophoresis*, **21**, 1145–1154.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K. and Weissman, J.S. (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737–741.

Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M.J., Cramer, R., Waterfield, M.D. and Timms, J.F. (2002) Evaluation of 2-D differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol. Cell. Proteomics*, **1**, 91–98.

Giege, P., Sweetlove, L.J. and Leaver, C.J. (2003) Identification of mitochondrial protein complexes in *Arabidopsis* using two-dimensional blue-native polyacrylamide gel electrophoresis. *Plant Mol. Biol. Rep.* **21**, 133–144.

Godovac-Zimmermann, J. and Brown, L.R. (2001) Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom. Rev.* **20**, 1–57.

Goff, S.A., Ricke, D., Lan, T.H. et al. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, **296**, 92–100.

Görg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruner, R. and Weiss, W. (2000) The current state of electrophoresis with immobilized pH gradients. *Electrophoresis*, **12**, 531–546.

Granier, F. (1988) Extraction of plant-proteins for two-dimensional electrophoresis. *Electrophoresis*, **9**, 712–718.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999.

Gygi, S.P., Rochon, Y., Franza, B.R. and Aebersold, R. (1999b) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730.

Hamdan, M. and Righetti, P.G. (2003) Assessment of protein expression by means of 2-D gel electrophoresis with and without mass spectrometry. *Mass Spec. Rev.* **22**, 272–284.

Handley, J. (2001) Proteomics systems emerge. *Anal. Chem.* **73**, 379A–383A.

Harry, J.L., Wilkins, M.R., Herbert, B.R., Packer, N.H., Gooley, A.A. and Williams, K.L. (2000) Proteomics: capacity versus utility. *Electrophoresis*, **21**, 1071–1081.

Hatzimanikatis, V. and Lee, K.H. (1999a) Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metab. Eng.* **1**, 275–281.

Hatzimanikatis, V., Choe, L.H. and Lee, K.H. (1999b) Proteomics: theoretical and experimental considerations. *Biotechnol. Prog.* **15**, 312–318.

Heazlewood, J.L. and Millar, A.H. (2003) Integrated plant proteomics – putting the green genomes to work. *Funct. Plant Biol.* **30**, 471–482.

Hoving, S., Voshol, H. and van Oostrum, J. (2000) Towards high performance two-dimensional gel electrophoresis using ultra-zoom gels. *Electrophoresis*, **21**, 2617–2621.

Hurkman, W.J. and Tanaka, C.K. (1986) Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. *Plant Physiol.* **81**, 802–806.

Ideker, T., Galitski, T. and Hood, L. (2001) A new approach to decoding life: systems biology. *Ann. Rev. Genom. Hum. Genet.* **2**, 343–372.

Jacobs, D.I., van Rijssen, M.S., van der Heijden, R. and Verpoorte, R. (2001) Sequential solubilization of proteins precipitated with trichloroacetic acid in acetone from cultured *Catharanthus roseus* cells yields 52% more spots after two-dimensional electrophoresis. *Proteomics*, **1**, 1345–1350.

Jung, J.W. and Lee, W. (2004) Structure-based functional discovery of proteins: structural proteomics. *J. Biochem. Mol. Biol.* **37**, 28–34.

Jung, E., Heller, M., Sanchez, J.C. and Hochstrasser, D. (2000) Proteomics meets cell biology: the establishment of subcellular proteomes. *Electrophoresis*, **21**, 3369–3377.

Kersten, B., Burkle, L., Kuhn, E.J., Giavalisco, P., Konthur, Z., Lueking, A., Walter, G., Eickhoff, H. and Schneider, U. (2002) Large-scale plant proteomics. *Plant Mol. Biol.* **48**, 133–141.

Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W. and Baginsky, S. (2004) The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. *Curr. Biol.* **14**, 354–362

Klose, J. (1999) Genotypes and phenotypes. *Electrophoresis*, **20**, 643–652.

Koller, A., Washburn, M.P., Lange, B.M. et al. (2002) Proteomic survey of metabolic pathways in rice. *Proc. Natl Acad. Sci. USA*, **99**, 11969–11974.

Komatsu, S., Muhammad, A. and Rakwal, R. (1999) Separation and characterization of proteins from green and etiolated shoots of rice (*Oryza sativa* L.): towards a rice proteome. *Electrophoresis*, **20**, 630–636.

Kruft, V., Eubel, H., Jansch, L., Werhahn, W. and Braun, H.P. (2001) Proteomic approach to identify novel mitochondrial proteins in *Arabidopsis. Plant Physiol.* **127**, 1694–1710.

Kubis, S., Baldwin, A., Patel, R., Razzaq, A., Dupree, P., Lilley, K., Kurth, J., Leister, D. and Jarvis, P. (2003) The *Arabidopsis ppi1* mutant is specifically defective in the expression, chloroplast import and accumulation of photosynthetic proteins. *Plant Cell*, **15**, 1859–1871.

Kuchler, M., Decker, S., Hormann, F., Soll, J. and Heins, L. (2002) Protein import into chloroplasts involves redox-regulated proteins. *EMBO J.* **21**, 6136–6145.

Lander, E.S., Linton, L.M., Biurren, B. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

Lauber, W.M., Carroll, J.A., Dufield, D.R., Kiesel, J.R., Radabaugh, M.R. and Malone, J.P. (2001) Mass spectrometry compatibility of two-dimensional gel protein stains. *Electrophoresis*, **22**, 906–918.

Lilley, K.S., Razzaq, A. and Dupree, P. (2002) Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr. Opin. Chem. Biol.* **6**, 46–50.

Lin, D., Tabb, D.L. and Yates, J.R. (2003) Large-scale protein identification using mass spectrometry. *Biochim. Biophys. Acta*, **646**, 1–10.

Lonosky, P.M., Zhang, X.S., Honavar, V.G., Dobbs, D.L., Fu, A. and Rodermel, S.R. (2004) A proteomic analysis of maize chloroplast biogenesis. *Plant Physiol.* **134**, 560–574.

Lopez, M.F. (2000) Better approaches to finding the needle in a haystack: optimizing proteome analysis through automation. *Electrophoresis*, **21**, 1082–1093.

Maltman, D.J., Simon, W.J., Wheeler, C.H., Dunn, M.J., Wait, R. and Slabas, A.R. (2002) Proteomic analysis of the endoplasmic reticulum from developing and germinating seed of castor (*Ricinus communis*). *Electrophoresis*, **23**, 626–639.

Manabe, T. (2000) Combination of electrophoretic techniques for comprehensive analysis of complex protein systems. *Electrophoresis*, **21**, 111–1122.

Mann, M. and Jensen, O.N, (2003) Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255–261.

Mann, M. and Wilm, M. (1995) Electrospray mass spectrometry for protein characterization. *Trends Biochem. Sci.* **20**, 219–223.

Mann, M., Hendrickson, R.C. and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Ann. Rev. Biochem.* **70**, 437–473.

Mathesius, U., Imin, N., Chen, H. *et al.* (2002) Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting. *Proteomics*, **2**, 1288–1303.

Medzihradszky, K.F., Campbell, J.M., Baldwin, M.A., Falick, A.M., Juhasz, P., Vestal, M.L. and Burlingame, A.L. (2000) The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **72**, 552–558.

Meyer, Y., Grosset, J., Chartier, Y. and Cleyet-Marel, J.C. (1988) Preparation by two-dimensional electrophoresis of proteins for antibody production: antibodies against proteins whose synthesis is reduced by auxin in tobacco mesophyll protoplasts. *Electrophoresis*, **9**, 704–712.

Mijnsbrugge, K.V., Meyermans, H., van Montagu, M., Bauw, G. and Boerjan, W. (2000) Wood formation in poplar: identification, characterization, and seasonal variation of xylem proteins. *Planta*, **210**, 589–598.

Millar, A.H., Sweetlove, L.J., Giege, P. and Leaver, C.J. (2001) Analysis of the *Arabidopsis* mitochondrial proteome. *Plant Physiol.* **127**, 1711–1727.

Miller, M.D., Acey, R.A., Lee, L.Y.-T. and Edwards, A.J. (2001) Digital imaging considerations for gel electrophoretic analysis systems. *Electrophoresis*, **22**, 791–800.

Molloy, M.P., Herbert, B.R., Walsh, B.J., Tyler, M.I., Trani, M., Sanchez, J.C., Hochstrasser, D.F., Williams, K.L. and Goodley, A.A. (1998) Extraction of membrane proteins by differential solubilization for separation using 2-D electrophoresis. *Electrophoresis*, **19**, 837–844.

Okushima, Y., Koizumi, N., Kusano, T. and Sano, H. (2000) Secreted proteins of tobacco cultured BY2 cells: identification of a new member of pathogenesis-related proteins. *Plant Mol. Biol.* **42**, 479–488.

Ong, S.E. and Pandey, A. (2001) An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol. Eng.* **18**, 195–205.

Patterson, S.D. (2004) How much of the proteome do we see with discovery-based proteomic methods and how much do we need to see? *Curr. Proteomics* **1**, 3–12.

Patterson, S.D. and Aebersold, R. (2003) Proteomics: the first decade and beyond. *Nat. Genet.* **33**, 311–323.

Patton, W.F. (2000a) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis*, **21**, 1123–1144.

Patton, W.F. (2000b) Making bind robots see: the synergy between fluorescent dyes and imaging devices in automated proteomics. *Biotechniques*, **28**, 944–957.

Peltier, J.B., Friso, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I. and van Wijk, K.J. (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. *Plant Cell*, **12**, 319–342.

Peltier, J.B., Emanuelsson, O., Kalume, D.E. *et al.* (2002) Central functions of the lumenal and peripheral thylakoid proteome of *Arabidopsis* determined by experimentation and genome-wide prediction. *Plant Cell*, **14**, 211–236.

Porubleva, L., Vander Velden, K., Kothari, S., Oliver, D.J. and Chitnis, P.R. (2001) The proteome of maize leaves: use of gene sequences and expressed sequence tag data for identification of proteins with peptide mass fingerprints. *Electrophoresis*, **22**, 1724–1738.

Puig, O., Caspary, F, Rigault, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M. and Seraphin, B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218–229.

Quadroni, M. and James, P. (1999) Proteomics and automation. *Electrophoresis*, **20**, 664–677.

Rabilloud, T. (1996) Solubilization of proteins for electrophoretic analyses. *Electrophoresis*, **17**, 813–829.

Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics*, **2**, 3–10.

Roberts, J.K. (2002) Proteomics and a future generation of plant molecular biologists. *Plant Mol. Biol.* **48**, 143–154.

Robertson, D., Mitchell, G.P., Gilroy, J.S., Gerrish, C., Bolwell, G.P. and Slabas, A.R. (1997) Differential extraction and protein sequencing reveals major differences in patterns of primary cell wall proteins from plants. *J. Biol. Chem.* **272**, 15841–15848.

Rohila, J.S., Chen, M., Cerny, R. and Fromm, M.E. (2004) Improved tandem affinity purification tag and methods for isolation of protein heterocomplexes from plants. *Plant J.* **38**, 172–181.

Rose, J.K.C., Catalá, C., Gonzalez-Carranza, Z.H. and Roberts, J. (2003) Plant cell wall disassembly. In *The Plant Cell Wall. Annual Plant Reviews Series* (Rose, J.K.C., ed.). London: Blackwell Publishing, pp. 264–324.

Rossignol, M. (2001) Analysis of the plant proteome. *Curr. Opin. Biotech.* **12**, 131–134.

Santoni, V., Bellini, C. and Caboche, M. (1994) Use of 2-dimensional protein-pattern analysis for the characterization of *Arabidopsis thaliana* mutants. *Planta*, **192**, 557–566.

Santoni, V., Doumas, P., Rouquié, D., Mansion, M. and Rabilloud, T. (1999) Large scale characterization of plant plasma membrane proteins. *Biochimie*, **81**, 655–661.

Santoni, V., Molloy, M. and Rabilloud, T. (2000a) Membrane proteins and proteomics: un amour impossible? *Electrophoresis*, **21**, 1054–1070.

Santoni, V., Kieffer, S., Desciaux, D., Masson, F. and Rabilloud, T. (2000b) Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis*, **21**, 3329–3344.

Saravanan, R.S. and Rose, J.K.C. (2004) A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics*, in press.

Schmid, M.B. (2002) Structural proteomics: the potential of high-throughput structure determination. *Trends Microbiol.* **10**, S27–S31.

Schuster, A.M. and Davies, E. (1983) Ribonucleic-acid and protein-metabolism in pea epicotyls.1. the aging process. *Plant Physiol.* **73**, 809–816.

Service, R.F. (2001) Proteomics. High-speed biologists search for gold in proteins. *Science*, **294**, 2074–2077.

Sherrier, D.J., Prime, T.A. and Dupree, P. (1999) Glycosylphosphatidylinositol-anchored cell-surface proteins from *Arabidopsis*. *Electrophoresis*, **20**, 2027–2035.

Standing, K.G. (2003) Peptide and protein de novo sequencing by mass spectrometry. *Curr. Opin. Struct. Biol.* **13**, 595–601.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Posgnan, F., Hawkins, E., Currie, I. and Davison, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*, **1**, 377–396.

Tsugita, A. and Kamo, M., (1999) 2-D electrophoresis of plant proteins. *Methods Mol. Biol.* **112**, 95–97.

Tsugita, A., Kamo, M., Kawakami, T. and Ohki, Y. (1996) Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. *Electrophoresis*, **17**, 855–865.

Unlu, M., Morgan, M.E. and Minden, J.S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*, **18**, 2071–2077.

Vener, A.V., Harms, A., Sussman, M.R. and Vierstra, R.D. (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. *J. Biol. Chem.* **9**, 6959–6966.

Venter, J.C., Adams, M.D., Myers, E.W. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.

Wang, W., Scali, M., Vignani, R. *et al.* (2003) Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis*, **24**, 2369–2375.

Washburn, M.P. and Yates, J.R. (2000) Analysis of the microbial proteome. *Curr. Opin. Microbiol.* **3**, 292–297.

Watson, B.S., Asirvatham, V.S., Wang, L. and Sumner, L.W. (2003) Mapping the proteome of barrel medic (*Medicago trunculata*). *Plant Physiol.* **131**, 1104–1123

Werhahn, W. and Braun, H.-P. (2002) Biochemical dissection of the mitochondrial proteome from *Arabidopsis thaliana* by three-dimensional gel electrophoresis. *Electrophoresis*, **23**, 640–646.

Whitelegge, J.P. (2002) Plant proteomics: BLASTing out of a Mud-PIT. *Proc. Natl Acad. Sci. USA*, **3**, 11564–11566.

van Wijk, K.J. (2001) Challenges and prospects of plant proteomics. *Plant Physiol.* **126**, 501–508.

Wildgruber, R., Hardre, A., Obermaier, C., Boguth, G., Weiss, W., Fey, S.J., Larsen, P.M. and Görg, A. (2000) Towards higher resolution: two-dimensional electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. *Electrophoresis*, **21**, 2610–2616.

Wilkins, M.R., Sanchez, J.C. and Williams, K.L. (1996) Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis*, **17**, 830–838.

Wilm, M. and Mann, M. (1996) Analytical properties of the nano-electrospray ion source. *Anal. Chem.* **68**, 1–8.

Woo, S.H., Fukuda, M., Islam, N., Takaoka, M., Kawasaki, H. and Hirano, H. (2002) Efficient peptide mapping and its application to identify embryo proteins in rice proteome analysis. *Electrophoresis*, **23**, 647–654.

Yan, J.X., Harry, R.A., Spibey, C. and Dunn, M.J. (2000) Post-electrophoretic staining of proteins separated by two-dimensional gel electrophoresis using SYPRO dyes. *Electrophoresis*, **21**, 3657–3665.

Yanagida, M. (2002) Functional proteomics; current achievements. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **5**, 89–106.

Yarmush, M.L. and Jayaraman, A. (2002) Advances in proteomic technologies. *Annu. Rev. Biomed. Eng.*, **4**, 349–373.

Yu, J., Hu, S., Wang, J. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, **296**, 79–92.

Zhou, G., Li, H., DeCamp, D. *et al.* (2002) 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol. Cell. Proteomics*, **1**, 117–124.

Zhu, H., Bilgin, M. and Snyder, M. (2003) Proteomics. *Ann. Rev. Biochem.* **72**, 783–812.